

# Thermal to Visible Face Recognition

Jonghyun Choi<sup>†</sup>, Shuowen Hu<sup>‡</sup>, S. Susan Young<sup>‡</sup> and Larry S. Davis<sup>†</sup>

<sup>†</sup>University of Maryland, College Park, MD

<sup>‡</sup>U.S. Army Research Laboratory, Adelphi, MD

## ABSTRACT

In low light conditions, visible light face identification is infeasible due to the lack of illumination. For nighttime surveillance, thermal imaging is commonly used because of the intrinsic emissivity of thermal radiation from the human body. However, matching thermal images of faces acquired at nighttime to the predominantly visible light face imagery in existing government databases and watch lists is a challenging task. The difficulty arises from the significant difference between the face's thermal signature and its visible signature (i.e. the modality gap). To match the thermal face to the visible face acquired by the two different modalities, we applied face recognition algorithms that reduce the modality gap in each step of face identification, from low-level analysis to machine learning techniques. Specifically, partial least squares-discriminant analysis (PLS-DA) based approaches were used to correlate the thermal face signatures to the visible face signatures, yielding a thermal-to-visible face identification rate of 49.9%. While this work makes progress for thermal-to-visible face recognition, more efforts need to be devoted to solving this difficult task. Successful development of a thermal-to-visible face recognition system would significantly enhance the Nation's nighttime surveillance capabilities.

**Keywords:** thermal, visible, thermal-to-visible face recognition, face, recognition, multi, modal

## 1. INTRODUCTION

For nighttime surveillance applications, acquisition of visible light imagery is impractical due to the lack of illumination. Thermal imaging, which acquires mid-wave infrared or long-wave infrared radiation naturally emitted by the human body, can be utilized in low light conditions to perform surveillance tasks. Identification of individuals captured by thermal imaging would significantly enhance nighttime intelligence gathering capabilities. However, government watch lists and databases almost exclusively contain visible light face imagery of individuals of interest. Matching thermal face imagery to the existing databases therefore requires the development of across modality face recognition algorithms and methods. Due to the large modality gap caused by the wavelength difference between visible and thermal radiation, thermal-to-visible face recognition is a challenging problem.

Face recognition has been an active area of research for the past two decades due its wide range of applications in law enforcement and verification/authentication systems. The focus of face recognition has primarily been on visible (located in the  $0.35\mu m$  to  $0.74\mu m$  wavelength range) imagery. Although much progress has been made, face recognition remains an open problem under uncontrolled lighting and pose conditions. More recently, some efforts have been devoted to face recognition using illumination invariant modalities such as infrared sensors.<sup>1-4</sup> The infrared spectrum consists of four main regions: near infrared (NIR;  $0.74-1\mu m$ ), shortwave infrared (SWIR;  $1-3\mu m$ ), mid-wave infrared (MWIR;  $3-5\mu m$ ), and long-wave infrared (LWIR;  $8-14\mu m$ ). While NIR and SWIR are also referred to as reflected infrared, MWIR and LWIR are naturally emitted by the human body and commonly referred to as thermal IR. Due to the proximity of the NIR spectrum to the visible spectrum, NIR face images preserves much of the information as in visible face images<sup>3,4</sup>.

Previous work<sup>1-3</sup> compared the performance of face recognition using NIR images with face recognition performance using visible images. The work of Lei and Li<sup>5</sup> and Klare and Jain<sup>6</sup> studied NIR-to-visible face recognition and the SWIR-to-visible face verification problem was addressed in the recent work of Bourlai *et al.*<sup>7</sup>

---

Further author information: (Send correspondence to Jonghyun Choi)

Jonghyun Choi: E-mail: jhchoi@umd.edu, Telephone: 1 301 335 3866

This work was done during the first author's internship at the U.S. Army Research Laboratory

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>APR 2012</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2012 to 00-00-2012</b>	
4. TITLE AND SUBTITLE <b>Thermal To Visible Face Recognition</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>U.S. Army Research Laboratory, Adelphi, MD, 20783</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>Preprint, presented at the SPIE Defense Security and Sensing Conference, 23 - 27 April 2012, Baltimore, MD</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>11</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

however, both NIR and SWIR require active illumination so it is not very practical for nighttime surveillance. The natural emission of thermal IR from the human body makes it an ideal modality for nighttime tasks, but the large disparateness between the thermal IR and visible spectrums results in a wide modality gap that makes thermal-to-visible face recognition a significantly more challenging problem than the NIR-to-visible or SWIR-to-visible face recognition problems. Although Xie *et al.*<sup>8</sup> have developed techniques for thermal-to-thermal face recognition and studies of Buysens and Revenu<sup>9</sup> have implemented fusion of thermal and visible images for face recognition, the authors have not found literature on thermal-to-visible face recognition.

The authors use as inspiration the following works that improved face recognition for visible imagery acquired under very different conditions such as illumination and pose. Algorithms developed for matching imagery acquired under very different conditions include correlating common discriminant analysis<sup>10</sup>, local metric learning methods<sup>5,11</sup>, and common subspace methods<sup>12</sup>. The most recent work of Sharma and Jacobs<sup>12</sup> proposed a general framework of doing multi-modal face recognition based on the existence of linear dependence between two modalities. We prove that the solution to the thermal-to-visible problem partially exists using similar arguments presented in the work of Sharma and Jacobs<sup>12</sup>.

The key to solving thermal-to-visible face recognition is the development of an algorithm or transform space that well-correlates the thermal and visible face signatures. Let  $R$  denote the physical face space, and  $P_V(\cdot)$  and  $P_T(\cdot)$  be the functions that map the physical face space to the visible and thermal face space, respectively. Then  $I_V = P_V(R)$  and  $I_T = P_T(R)$  represent the visible and thermal face as acquired by the visible and thermal sensor, as illustrated in Figure 1. If  $I_V$  intersects with  $I_T$ , then thermal-to-visible face recognition is feasible.

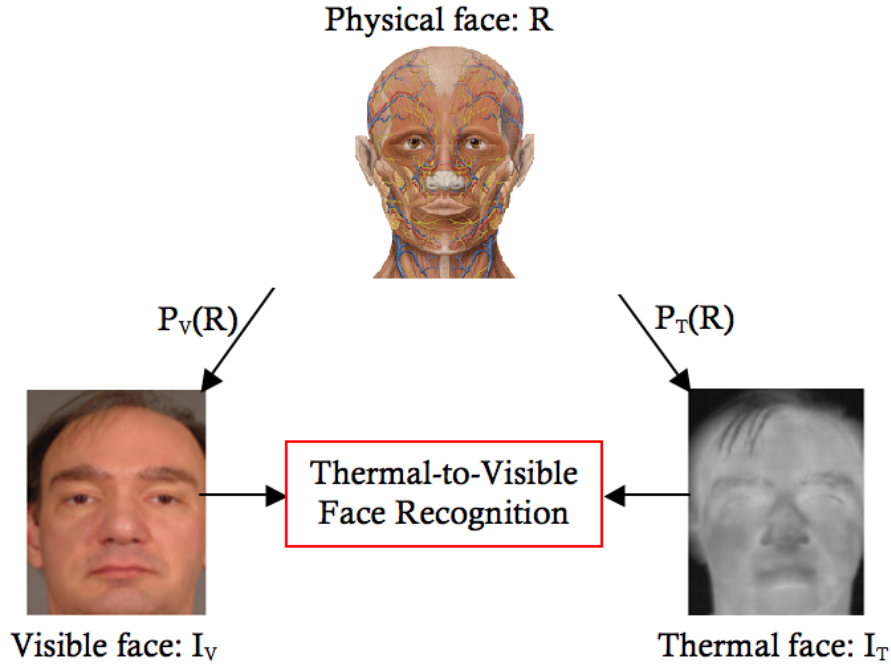


Figure 1. Illustration of visible and thermal representations of the physical face

In this work, we study the problem of matching thermal probe images to visible gallery images. The gallery imagery consists of visible images to simulate government watch lists, and the thermal IR probe imagery simulates suspect imagery acquired during nighttime surveillance operations. We cast this face identification problem of matching thermal probe images to visible gallery images as a multi-modal face recognition problem. Although there are several previous studies dealing with across modality NIR-to-visible face recognition<sup>5,11</sup>, to the best of our knowledge, this work is the first in trying to match thermal face images to visible face images.

To tackle this problem, we explore various pre-processing techniques such as self-quotient images<sup>13</sup> and difference-of-Gaussian<sup>14</sup> filtering as well as various feature transforms to reduce the variations in each domain

and enhance the multi-modal matching. In addition, we make use of a discriminant modeling function to weight the feature vectors by maximizing covariance between two modalities using partial least squares (PLS) analysis. We also applied the recently proposed multi-modal face recognition technique of common subspace construction<sup>12</sup> for comparison.

## 2. PRE-PROCESSING

Since thermal and visible face images have very different signatures, preprocessing is important in solving the thermal-to-visible face recognition problem. For this work, preprocessing consists of two main stages: thermal image normalization, and local variation reduction for thermal and visible imagery. Note that the dead pixels within the thermal imagery were removed via simple median filtering prior to image normalization.

As a first pre-processing step for thermal imagery, we normalize the thermal signatures by its mean and standard deviation to reduce the temperature offset and statistical variation across thermal images. Equation (1) is the normalization equation (parameter  $\alpha$  was set to 5 for this work).

$$Z(x, y) = \frac{I(x, y) - \hat{I}}{\alpha\sigma} \quad (1)$$

where  $(x, y)$  are image coordinates,  $Z(x, y)$  is the normalized intensity value at  $(x, y)$ ,  $I(x, y)$  is the intensity value at  $(x, y)$ ,  $\hat{I}$  is the mean of all intensity values in the image, and  $\sigma$  is the standard deviation of intensity values in the image. Fig. 2 shows (a) the original thermal face image and (b) the normalized thermal image warped to canonical position.



Figure 2. (a) Original thermal image (b) Normalized thermal image warped to canonical position.

The second preprocessing step adjusts the thermal and visible imagery for local variations. For visible imagery, illumination primarily induces the local variations, whereas for the thermal imagery, the varying heat distribution within the face produces the local variations. Self quotient image (SQI) and difference of Gaussian filtering (DOG) have been commonly applied to reduce illumination variations in visible face imagery. They were also applied here to reduce the local variations in thermal face imagery. Fig. 3 shows two pre-processed images along with the original images in the visible and thermal domains. As can be observed, SQI emphasizes the edge information in the thermal imagery while DOG filtering blurs the visible imagery.

## 3. FEATURE TRANSFORMS

Selecting good features is very important in computer vision applications<sup>15</sup>. Many feature descriptors have been developed to facilitate face recognition, such as local binary patterns (LBP)<sup>16</sup>, its multi-scale variants (MSLBP)<sup>17</sup>, and Gabor filter<sup>18</sup>. Recently, Schwartz *et al.*<sup>19</sup> proposed using large set of combined features to improve performance.

Local binary patterns (LBP) is a well-known texture descriptor and a successful local descriptor for face recognition under local illumination variations<sup>16,20</sup>. LBP descriptors are compact and easy to compare by various histogram metrics. In addition, there are many LBP variants to improve the description performance of LBP. The most popular extension is multi-scale LBP (MSLBP) which uses multiple radii<sup>17</sup>.

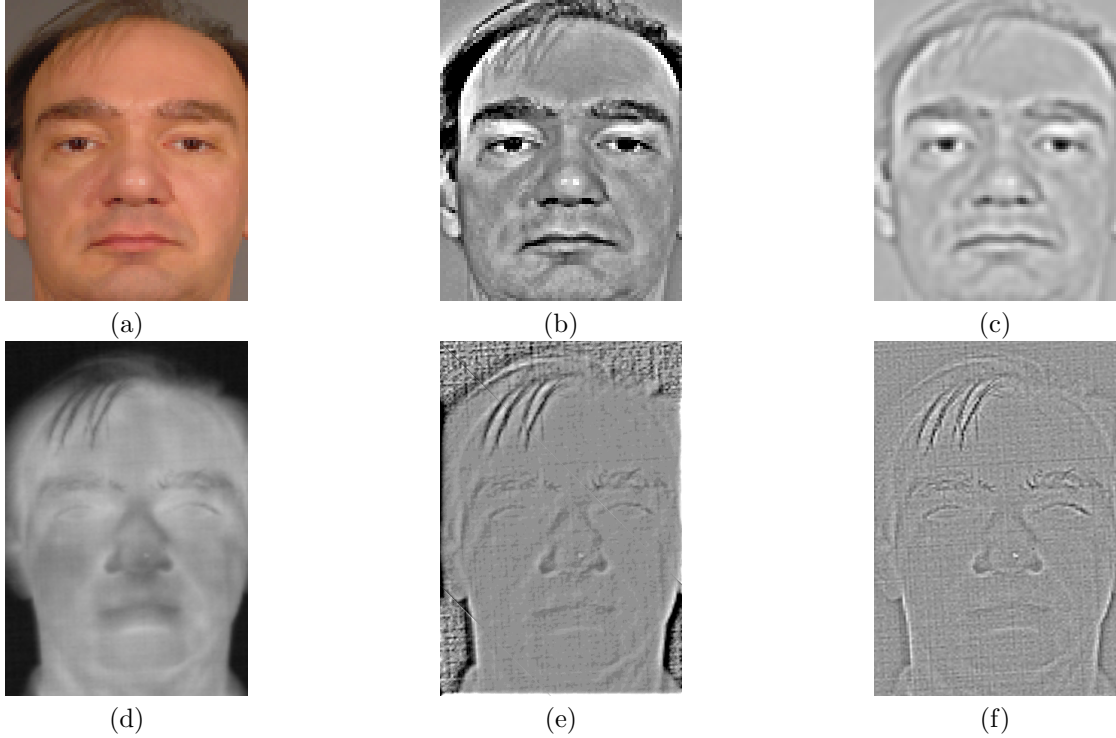


Figure 3. Effect of Pre-Processing of Local Variations. (a) Original visible image (b) SQI applied to visible image (c) DOG filter applied to visible image (d) Original thermal image (e) SQI applied to thermal image (f) DOG filter applied to thermal image

The histogram of oriented gradients (HOG) has been successfully applied to tasks such as human detection<sup>21</sup> and face recognition<sup>22</sup>. Similar to LBP, edge information captured by gradients within blocks is packed into a histogram. Discarding pixel location information by block-based histogram binning, LBP and HOG gain invariance to local changes such as small facial expressions and pose variations in pedestrian images.

The Gabor wavelets are also effective face descriptors which capture global shape information centered on a pixel<sup>18</sup>. The convolution of multiple Gaussian-like kernels at different scales and orientations captures information insensitive to expression variation and blur at a pixel's location.

We consider all these features for thermal-to-visible face recognition. We also compare the results of using raw intensity values as a feature as in some previous works<sup>12, 23</sup>.

#### 4. PLS-DA BASED FACE RECOGNITION

Given a feature vector obtained from pre-processing and feature transforms, we formulate the extreme multi-modal face recognition problem via partial least squares-discriminant analysis (PLS-DA) face recognition framework. We compare the results of PLS-DA approach with that of a PLS-based common subspace face recognition method.

The pre-processing and feature transform stages (Sections 2 and 3) reduce the modality gap between the two sensors. We build a discriminant PLS-DA model to further reduce the modality gap and classify the probe more discriminatively without additional training set<sup>22</sup>.

Once we built discriminant model for each subject in the visible gallery, we evaluate the thermal probe images by simply taking the dot product of extracted features from the images to build models. This process is very fast and easily parallelized. The subject whose response is the maximum among gallery images is selected as the identified subject. Fig. 4 illustrates the training and testing procedure.

Partial least squares analysis is recently gaining popularity in the computer vision community as a supervised dimension reduction technique specifically useful in multi-collinearity situations<sup>12, 22, 24, 25</sup>. Multi-collinearity is a phenomenon of high dependence between variables (feature dimension), which happens very frequently with large dimensional features. Modern computer vision algorithms take advantage of many features to capture different kinds of low level information. Consequently, the resulting feature vectors usually have large dimension and suffer from multi-collinearity. PLS-DA is a PLS regression based discriminant analysis and is known to perform well in multi-collinear situations<sup>26</sup>.

#### 4.1 Model Building Stage

The main idea of the PLS regression is to maximize the covariance between the dependent variable  $Y$  and a weighted sum of the independent variables in  $X$  by finding a weighting vector  $w$  as in Eq. (2).

$$\text{cov}(t, u)^2 = \max_{|w|=1} \text{cov}(Xw, Y)^2 \quad (2)$$

where  $t$  and  $u$  are the column vectors of matrices  $T$  and  $U$  to be described in Eq. (3), respectively, and  $\text{cov}(t, u)$  is the covariance matrix between latent vectors  $t$  and  $u$ .

In our application,  $X$  corresponds to a feature vector and  $Y$  corresponds to a label which is defined by 1 or 0 for one-vs-all scheme.  $Y$  is a binary vector whose  $i^{\text{th}}$  element is 1 if the  $i^{\text{th}}$  row of  $X$  belongs to the given subject for the subject's model. For our work,  $X$  is the set of visible subject images in the gallery, and  $Y$  is the corresponding one-vs-all subject labels; no thermal imagery was used during the model building stage for PLS-DA. Next, we explain how the weighting vector is obtained by PLS in more detail.

Let  $\mathcal{X} \subset \mathbb{R}^m$  denote an  $m$ -dimensional feature space,  $\mathcal{Y} \subset \mathbb{R}$  denote a scalar space representing the response variable.  $X$  and  $Y$  are normalized by subtracting its average across the columns. PLS decomposes a  $(n \times m)$  matrix  $X \in \mathcal{X}$  (where  $n$  denotes the number of samples) and the  $(n \times 1)$  vector  $Y \in \mathcal{Y}$  into

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned} \quad (3)$$

The  $(n \times p)$  matrices  $T$  and  $U$  are called *scores* that contain latent vectors, the  $(m \times p)$  matrix  $P$  and the  $(1 \times p)$  vector  $Q$  are called *loadings*, and the  $(n \times m)$  matrix  $E$  and the  $(n \times 1)$  vector  $F$  are called *residuals*. Using a greedy algorithm called NIPALS<sup>27</sup>, we can obtain a set of weight vectors iteratively, stored in the matrix  $W = (w_1, w_2, \dots, w_p)$ . At the end of each iteration, the matrices  $X$  and  $Y$  are deflated by subtracting their rank-one approximations based on  $t$  and  $u$  and this is continued until the desired number of latent vectors is obtained, denoted by  $p$ . More detailed procedure is explained in Rosipal *et al.*<sup>26</sup>.

Once we obtain  $W$ , we can compute the regression vector from  $X$  to  $Y$  using the following equations,

$$\begin{aligned} Y &= XB + F^* \\ B &= W(P^TW)^{-1} = X^TU(T^TX X^TU)^{-1}Y \end{aligned} \quad (4)$$

where  $B$  is the regression vector and  $F^*$  is a residual vector. We refer to the regression vector as the *model*.

An additional advantage of using PLS-DA is that it works well with extremely unbalanced data, *i.e.* number of positive samples is 1 while number of negative samples is large, say 7,000<sup>22, 25</sup>. In our protocol, we use many positive samples (each subject has multiple images) so that PLS-DA model is stable. Moreover, one additional advantage of using the PLS-DA approach is that a separate training set is not required. It only uses the gallery images to build a discriminant model for each subject in the gallery.

## 4.2 Testing Stage

In the testing stage, we pre-process the input thermal image and perform feature extraction using the same features as for the gallery. The probe (feature vector extracted from thermal face image) must be first centered and normalized. Since probes are inputted one by one, it is not possible to obtain an estimate of the probe's true underlying mean and standard deviation. A common practice is to normalize the probe using statistics estimated from the gallery. For this work, the mean and standard deviation estimated from the feature vectors of the visible gallery are therefore used to center and normalize each thermal probe feature vector, under the assumption that the pre-processing and feature transformation have reduced the modality gap. Then we take a single dot product of the feature vector from the thermal probe image with the models obtained from model building stage. Eq. (5) and (6) define the testing stage:

$$\hat{Y}_i = \frac{X_q - \bar{X}_t}{\sigma_t} \cdot B_i, \quad \forall i \in S_{TR} \quad (5)$$

$$\hat{X}_q = \arg \max_i \hat{Y}_i \quad (6)$$

where  $\hat{Y}_i$  denotes response of  $i^{th}$  model,  $B_i$  denotes  $i^{th}$  model,  $X_q$  denotes thermal probe feature vector,  $\bar{X}_t$  denotes a sample mean of the visible model building set,  $\sigma_t$  denotes a standard deviation of training set,  $S_{TR}$  denotes training set,  $\hat{X}_q$  denotes the identified result for the probe image and division of vector refers to dimension wise division.

PLS-DA model is a linear feature weighting and the resulting scalar value from the dot product, called *response* as depicted in Fig. 4, is used to find the closest face in the gallery by choosing the face that has the maximum response. This approach is scalable to size of the gallery. When the number of gallery subjects/images increases, the number of dot products increases linearly - the testing time only increases linearly. Since each one-vs-all process is independent, this algorithm is readily parallelized so that it is scalable to large gallery sizes.

## 5. DATASET

We used the thermal and visible dataset (X1 Collection) from the University of Norte-Dame (UND) for this work, specifically the part of the dataset containing 82 subjects with multiple thermal and visible images for each subject. We partitioned the 82 subjects (2,292 images total) into two sets of 41 subjects each, call them Set A and Set B.

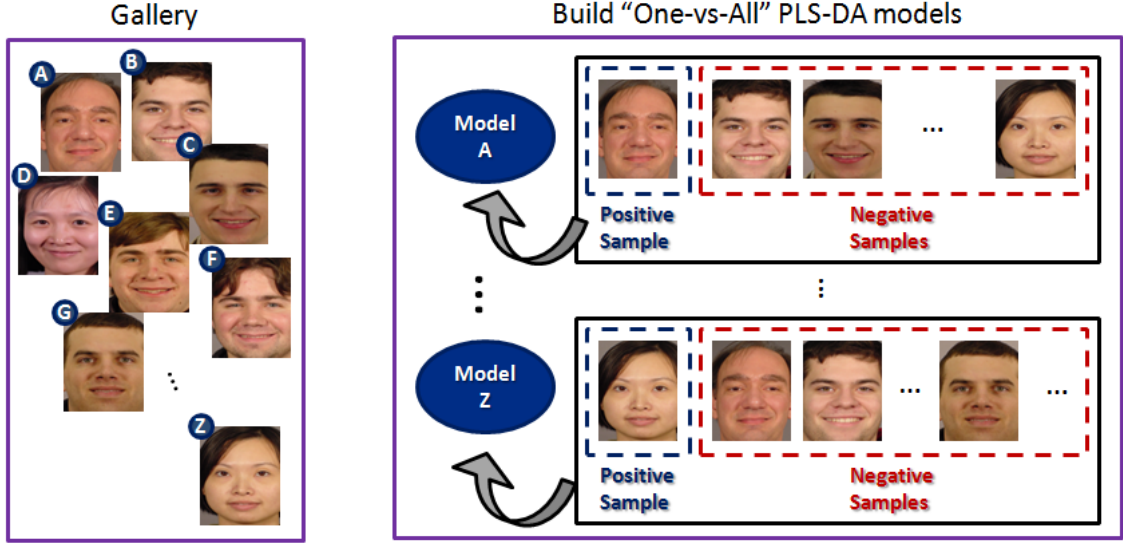
Since the PLS-based common subspace approach requires training, Set A was used for training and Set B was used for testing. Note that the visible imagery of Set B was used as the gallery and thus projected in advance as a whole, whereas the thermal images of Set B were used as probes and therefore projected one by one to find the nearest neighbor within the projected gallery.

In contrast to the PLS-based common subspace approach, PLS-DA does not require a separate training set. The visible imagery of Set B was used as the gallery for the one-vs-all model building for PLS-DA, and the thermal imagery of Set B was used as the probe during testing. The partition into two sets was done so that the PLS-DA results can be compared fairly with the PLS-based common subspace results (*i.e.* so that number of subjects/images was the same for both approaches).

## 6. EXPERIMENTAL RESULTS AND DISCUSSION

We present here the experimental results for the UND dataset. We performed basic pre-processing consisting of dead pixel removal, affinely warping the face by four fiducial points (two eyes, nose tip, mouth), cropping to face regions, and resizing to 80×100 pixels. We performed thermal-to-visible face recognition with PLS-DA, and compared the results to that of a PLS based common subspace method.

## Model Building (Training)



## Testing

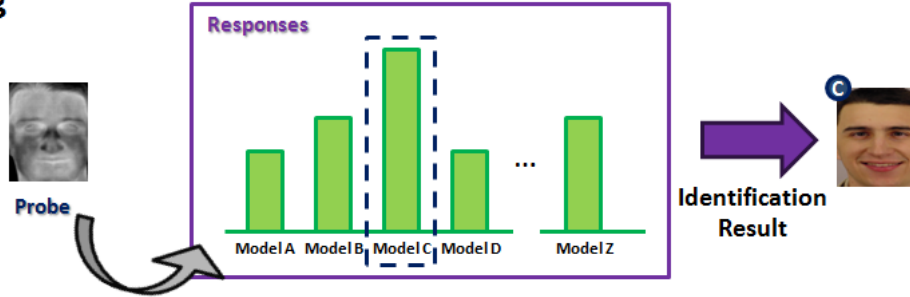


Figure 4. Illustration of PLS-DA Face Recognition Framework

### 6.1 PLS-DA Face Recognition

For the experiments of PLS-DA framework, we build a discriminant model for each gallery face with one-against-all scheme. Once we build the model, the thermal probe image is evaluated by dot product to each model of gallery to get the similarity responses. The subject whose response is the maximum among the gallery images is selected as identified subject.

Even though visible and thermal images have different visual signatures, appropriate pre-processing techniques and feature transforms can produce consistent low level signature and therefore reduce the modality gap. With narrowed modality gap by pre-processing, the PLS-DA with a set of features is expected to perform well. For feature transform, we used different kinds of features to capture diverse information (HOG, LBP, MSLBP and Gabor), followed by PLS regression weighting by one-against-all methods for matching the thermal probe to the visible gallery sets<sup>19</sup>.

#### 6.1.1 Different Pre-Processing Schemes and Feature Transforms

We investigated different kinds of pre-processing and feature transform schemes. Table 1 shows the recognition rate for each combination of preprocessing scheme and feature transform.

Different pre-processing schemes give different low level information to PLS-DA models. We explored SQI and DOG filtering, and found DOG is qualitatively and empirically better than SQI. Since DOG blurs the image and given the fact that the thermal image is inherently smoother, DOG filtering reduces the modality gap more effectively. Different feature transforms interpret the image intensity in different manner but there is *no free*



*lunch* in feature transform<sup>28,29</sup>, which means there is no universally good feature. Since the performance of feature transforms depends on the subsequent machine learning stage, we compare the effectiveness of HOG, LBP, MSLBP and Gabor features for PLS-DA, and choose the feature that produces the best identification rate.

The best combination is DOG filtering and HOG features. The reason that HOG with DOG performs the best is that DOG makes the images spatially smooth so the gradient information becomes more stable. On the other hand, LBP is sensitive to subtle pixel-wise differences, which was lost due to the spatial smoothing during pre-processing. Since Gabor feature is a response of non-isotropic Gaussian kernel, the improvement by preprocessing is expected to be lower than HOG or LBP. Table 1 tabulates the rank-1 face identification results, showing that the results of Gabor with different pre-processing do not improve much, whereas the results of HOG and LBP significantly improve with pre-processing.

Table 1. Rank-1 Face Identification Rate (%) of PLS-regression based approach (PLS-DA) with different pre-processing schemes and different feature transforms.

Features Pre-Processing	LBP	MSLBP	HOG	Gabor
None	7.7	8.3	17.3	34.9
SQI	14.7	14.3	39.8	19.4
DOG	26.1	22.7	<b>49.9</b>	35.7

## 6.2 Comparison to PLS based Common Subspace Approach

We compared the performance of the PLS-DA approach with that of the state-of-the-art multi-modal PLS-based common subspace face recognition technique. Although the common subspace method by PLS<sup>12</sup> for our problem is expected to perform poorly, there is hope that edges can correlate both modalities. The rank-1 face identification rate of PLS based common subspace method is 10.6% (using intensity values as features). Although the PLS based common subspace framework is generalized for any multi-modal situation, it is limited to the case that the solution of their linear regression exists.

**Partial-Existence of Solution** Consider the equation of two images from each modality as following

$$I_v = TP_v R_k, \quad I_t = TP_t R_k \quad (7)$$

where  $R_k$  denotes original face,  $P_v$  and  $P_t$  denote camera projection by visible camera and thermal camera respectively,  $T$  is a feature transform and  $I_v$  and  $I_t$  are visible and thermal images respectively.

If  $I_v$  and  $I_t$  can be modeled in a subspace by PLS weighting vectors for each modality, denoted by  $w$  and  $c$ , the following equation should be satisfied

$$\begin{aligned} w^T I_v &= c^T I_t \\ w^T TP_v R_k &= c^T TP_t R_k \\ w^T TP_v &= c^T TP_t. \end{aligned} \quad (8)$$

If and only if  $TP_v$  and  $TP_t$  intersect each other in a feature space generated by a transform  $T$ , the solution  $w^T$  and  $c^T$  exists and the multi-modal face recognition can be solved by PLS modeling. But in our problem,  $P_v$  and  $P_t$  are from different sensors so there is little intensity-wise correlation other than the main edges. Therefore the solution obtained by this approach is not accurate since  $P_v$  and  $P_t$  are not linearly well-correlated.

It is of interest to see if we can use different feature transforms other than the intensity values to improve classifier performance since feature transforms usually give more discriminative information. We tested Gabor

features for the PLS based common subspace approach. Although Gabor transforms is expected to describe the face more meaningfully than intensity features for the purposes of classification, however, face recognition performance using Gabor features is 6.36%, which is surprisingly less than the performance using intensity features. A possible reason of poor performance by Gabor feature is that the Gabor transform distorts the linear relationship between the thermal and visible modalities so that it gets more difficult to linearly correlate information coming from two modalities. The pre-processing is not applied for the common subspace method since the common subspace method explicitly model the modality gap and pre-processing might be harmful for the linear relationship. In some preliminary experiments, the common subspace approach with pre-processing yielded significantly worse results than without preprocessing.

Comparing the two methods in Table 2, we show that the PLS-DA with pre-processing scheme performs better than the PLS subspace method. The most likely reason that PLS-DA is better is that it explicitly models discriminative features of the galleries images whereas the PLS-based common subspace method fails to.

Table 2. The best Rank-1 Face Identification Rate (%) of different approaches

Methods	Rate(%)
PLS-DA	<b>49.9</b>
PLS Subspace <sup>12</sup>	10.6

## 7. CONCLUSION AND FUTURE WORK

In this study, we investigated the thermal-to-visible face recognition problem, which has a wide modality gap. We showed that our novel combination of pre-processing, feature transforms, and PLS-DA recognition framework can achieve 49.9% accuracy, which is almost 40% better than the performance of the PLS-based common subspace face recognition approach. As future work, we will investigate the modality gap more explicitly, *e.g.*, using learning based pre-processing method. Development of an effective thermal-to-visible face recognition system is expected to significantly improve nighttime surveillance capabilities.

## REFERENCES

- [1] Chen, X., Flynn, P. J., and Bowyer, K. W., “IR and visible light face recognition,” *CVIU* **99**(3), 332–358 (2005).
- [2] Kong, S. G., Heo, J., Abidi, B. R., Paik, J., and Abidi, M. A., “Recent advances in visual and infrared face recognition: a review,” *CVIU* **97**, 103–135 (2005).
- [3] Li, S. Z., Chu, R., Liao, S., and Zhang, L., “Illumination Invariant Face Recognition Using Near-Infrared Images,” *IEEE T. PAMI* **29**(4), 627–639 (2007).
- [4] Buyssens, P. and Revenu, M., “IR and visible face identification via sparse representation,” in *[BTAS]*, 1–6 (2010).
- [5] Lei, Z. and Li, S. Z., “Coupled Spectral Regression for matching heterogeneous faces,” in *[CVPR]*, (2009).
- [6] Klare, B. and Jain, A. K., “Heterogeneous Face Recognition: Matching NIR to Visible Light Images,” in *[ICPR]*, (2010).
- [7] “Cross-spectral Face Verification in the Short Wave Infrared (SWIR) Band,” in *[ICPR]*, (2010).
- [8] Xie, Z., Liu, G., Wu, S., and Fang, Z., “Infrared Face Recognition Based on Blood Perfusion and Fisher Linear Discrimination Analysis,” in *[International Workshop on Imaging Systems and Techniques]*, (2009).
- [9] Buyssens, P. and Revenu, M., “Fusion levels of visible and infrared modalities for face recognition,” in *[BTAS]*, 1–6 (2010).
- [10] Lin, D. and Tang, X., “Inter-modality Face Recognition,” in *[ECCV (4)]*, 13–26 (2006).
- [11] Zhang, W., Wang, X., and Tang, X., “Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition,” in *[CVPR]*, (2011).
- [12] Sharma, A. and Jacobs, D., “Bypassing Synthesis: PLS for Face Recognition with Pose, Low-Resolution and Sketch,” in *[CVPR]*, 593–600 (2011).

- [13] Wang, H., Li, S. Z., and Wang, Y., “Face recognition under varying lighting conditions using self quotient image,” in *[FG]*, 819–824 (2004).
- [14] Tan, X. and Triggs, B., “Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions,” *IP* **19**(6), 1635–1650 (2010).
- [15] Vedaldi, A., Ling, H., and Soatto, S., “Knowing a Good Feature When You See It: Ground Truth and Methodology to Evaluate Local Features for Recognition,” in *[Computer Vision: Detection, Recognition and Reconstruction]*, Cipolla, R., Battiato, S., and Farinella, G. M., eds., **285**, 27–49 (2010).
- [16] Ahonen, T., Hadid, A., and Pietikainen, M., “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE T. PAMI* **28**(12), 2037–2041 (2006).
- [17] Ojala, T., Pietikainen, M., and Maenpaa, T., “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *PAMI* **24**(7), 971–987 (2002).
- [18] Shen, L. and Bai, L., “A review on Gabor wavelets for face recognition,” *Pattern Anal. Appl.* **9**, 273–292 (2006).
- [19] Schwartz, W. R., Guo, H., Choi, J., and Davis, L. S., “Face Identification Using Large Feature Sets,” *IEEE Trans. on Image Processing* (To appear).
- [20] Heikkilä, M., Pietikäinen, M., and Schmid, C., “Description of interest regions with local binary patterns,” *Pattern Recognition* **42**, 425–436 (2009).
- [21] Dalal, N. and Triggs, B., “Histograms of Oriented Gradients for Human Detection,” in *[CVPR]*, (2005).
- [22] Schwartz, W. R., Guo, H., and Davis, L. S., “A Robust and Scalable Approach to Face Identification,” in *[ECCV]*, (2010).
- [23] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y., “Robust Face Recognition via Sparse Representation,” *IEEE Trans. PAMI* **31**(2), 210–227 (2009).
- [24] Kembhavi, A., Harwood, D., and Davis, L. S., “Vehicle Detection Using Partial Least Squares,” *IEEE T. PAMI* **99** (2010).
- [25] Schwartz, W. R., Kembhavi, A., Harwood, D., and Davis, L. S., “Human detection using partial least squares analysis,” in *[ICCV]*, 24–31 (2009).
- [26] Rosipal, R. and Krämer, N., “Overview and Recent Advances in Partial Least Squares,” in *[Subspace, Latent Structure and Feature Selection Techniques]*, (2006).
- [27] Wold, H., *[Partial Least Squares]*, vol. 6, 581–591 (1985).
- [28] Wolpert, D. H. and Macready, W. G., “No free lunch theorems for optimization,” *Evolutionary Computation, IEEE Transactions on* **1**(1), 67–82 (1997).
- [29] Mikolajczyk, K. and Schmid, C., “A performance evaluation of local descriptors,” *IEEE T. PAMI* **27**(10), 1615–1630 (2005).